applied sciences

MDPI

Article

Proactive Artificial Intelligence: Evaluating Prompt Timing in Autonomous Driving Contexts

Simone Piersigilli [1] **and Giandomenico Caruso** [2,*]

[1] Department of Design, Politecnico di Milano, 20158 Milan, Italy; simone.piersigilli@polimi.it
[2] Department of Mechanical Engineering, Politecnico di Milano, 20156 Milan, Italy
* Correspondence: giandomenico.caruso@polimi.it; Tel.: +39-2399-8094

Featured Application

This work supports the design of timing-aware proactive Artificial Intelligence in intelligent cockpits, helping systems deliver prompts at moments that improve trust, usefulness, and user experience while limiting cognitive load. The proposed framework can guide in-vehicle assistants in providing anticipatory, real-time, and reflective support for tasks such as navigation, safety, and infotainment, and can be extended to other domains requiring well-timed human–machine interaction.

Abstract

Proactive Artificial Intelligence systems in intelligent cockpits can initiate prompts without explicit user commands, yet when such prompts should be delivered remains underexplored. This study examines how prompt timing affects user experience in autonomous driving contexts. Using a Virtual prototype developed in Unity and deployed on Meta Quest, 28 participants experienced a tourism-oriented autonomous driving scenario in a within-subjects design, encountering proactive prompts at three temporal positions relative to driving events: Before, During, and After. User experience was assessed across four dimensions using validated scales. Repeated-measures ANOVA revealed significant effects of prompt timing on all measures ($p < 0.001$, $\eta^2p = 0.35$–0.59). Prompts delivered before and during events were consistently rated higher than those delivered after, particularly for trust, usefulness, and satisfaction. Differences between Before and During conditions were limited to overall experience satisfaction, while During prompts were associated with higher cognitive load. These findings suggest that temporal alignment between system behavior and user cognitive processes plays a key role in shaping interaction quality. A layered timing framework is proposed that assigns anticipatory, real-time, and reflective functions to prompts before, during, and after, respectively. Further studies in real-world contexts are needed to validate these results.

Keywords: proactive artificial intelligence; intelligent cockpits; prompt timing; temporal alignment; human–machine interaction; autonomous driving; user experience; cognitive load; trust

Academic Editors: Tomislav Stipančić, Katerina Kabassi and Yuchong Zhang

Received: 20 April 2026
Revised: 29 May 2026
Accepted: 1 June 2026
Published: 8 June 2026

1. Introduction

Intelligent cockpits evolve from interface-centered information systems into cooperative environments in which Artificial Intelligence (AI) agents perceive context, anticipate user needs, and initiate interaction without explicit commands. In this transition, the design problem is no longer limited to what the system should communicate but extends

to when communication should occur. Buyukgoz et al. define proactive AI as the ability to "autonomously initiate anticipatory action based on reasoning" [1], while recent intelligent cockpit research has emphasized context awareness, multimodal integration, and adaptive interaction as key enablers of next-generation in-vehicle experience [2]. Yet, increased proactivity does not automatically translate into better cooperation. If system interventions are poorly timed, they can be experienced as disruptive rather than intelligent, thereby undermining trust, perceived usefulness, and the overall quality of interaction [3].

Across human–computer interaction research, prompt timing has emerged as a consequential design variable rather than a secondary implementation detail. Prior studies on interruptions, notifications, and adaptive interventions show that the effectiveness of system prompts depends on their relationship to task structure, attentional availability, and user receptivity [4–6]. In particular, timing frameworks based on task phases, interruptibility windows, and just-in-time support suggest that interventions are best received when they align with the user's cognitive rhythm rather than with system convenience alone [4–6]. This issue is especially salient in vehicle environments, where the appropriateness of an intervention depends on rapidly changing attentional demands and situational conditions [7,8]. However, most automotive work has focused on operator-centered scenarios such as warnings, take-over requests, and navigation support. As automated driving advances toward Level 3 and beyond, the user's role progressively shifts from continuous vehicle control to supervision, interpretation, and experience-oriented engagement. The timing of proactive AI prompts in these more cooperative, experience-centered cockpit conditions remains insufficiently understood [2,8].

This gap is particularly consequential in tourism-oriented autonomous driving contexts. Unlike routine commuting, tourism travel involves frequent shifts among observation, planning, environmental interpretation, and spontaneous decision-making, creating natural fluctuations in interruptibility and cognitive readiness [8,9]. In such contexts, proactive prompts are not merely functional notifications; they become part of the journey experience itself. The design challenge is therefore not only to make proactive AI informative but also to make it temporally appropriate within a cooperative human–machine interaction loop. A prompt delivered too late may signal weak situational awareness; a prompt delivered during an inopportune moment may increase cognitive burden; and even useful content may be rejected if it fails to match the user's temporal state [3–7].

Against this background, this paper investigates how proactive AI prompt timing shapes user experience in an intelligent cockpit scenario for autonomous tourism driving. Rather than treating timing as a minor interface parameter, the study positions it as a core variable in cooperative interaction design. In a within-subjects experiment conducted in a Virtual Reality (VR) environment, the research compares identical prompts delivered before, during, and after salient driving events and evaluates their effects on cognitive load, trust in automation, perceived usefulness, and overall user experience. The paper's contribution is twofold and operates on two distinct levels. Empirically, it demonstrates that prompt timing, when isolated from content and modality, produces systematic and simultaneous effects across cognitive load, trust, perceived usefulness, and overall experience in an intelligent cockpit scenario. Conceptually, it advances temporal alignment as the explanatory construct that organizes these effects, defining it as the degree to which system intervention synchronizes with the user's cognitive rhythm. The empirical contribution establishes that timing matters and how it matters; the conceptual contribution proposes a framework through which these effects can be interpreted and translated into design logic. The two contributions are complementary rather than redundant: the empirical findings give the construct its anchor, while the construct gives the findings explanatory coherence. This contribution matters because it reframes proactive intelligence in coopera-

tive human–machine interaction not as mere responsiveness, but as temporal sensitivity. On that basis, the paper argues that intelligent cockpit systems should move beyond fixed or event-triggered prompting toward timing-aware cooperative support, and it derives a layered design logic for anticipatory, concurrent, and reflective prompting in autonomous travel contexts.

The remainder of this paper is organized as follows. Section 2 reviews related work and identifies the research gap this study addresses. Section 3 introduces the conceptual framework and formulates the research hypotheses. Section 4 describes the methodology, including the study design, experimental setup, protocol, and data analysis procedures. Section 5 presents the results, covering both descriptive findings and inferential analyses across all measured dimensions. Section 6 discusses the implications of the results in relation to prior work and outlines key insights. Finally, Section 7 concludes the paper and highlights directions for future research.

2. Related Work and Research Gap

Proactive AI has been defined not simply as system initiative, but as the capacity to anticipate, reason, and act autonomously in ways that affect users and their environments [1]. In human–computer interaction (HCI), this shift from reactive response to proactive intervention has been accompanied by a parallel shift in the design problem itself: the issue is no longer only what information a system should provide but also how initiative should be orchestrated so that it is perceived as useful rather than intrusive. Recent work on proactive AI systems and dialog strategies shows that user acceptance depends heavily on whether system behavior appears contextually appropriate, trustworthy, and well-judged in relation to ongoing activity [3,10]. This is especially relevant for intelligent cockpits, where AI is increasingly expected to operate as a collaborative agent rather than as a passive interface layer [2].

Existing research offers several complementary frameworks for understanding when proactive systems should intervene. A first stream models timing in relation to task structure, distinguishing interventions that occur before, during, or after task execution and showing that interruptions are better tolerated near task boundaries than at cognitively dense moments [4]. A second stream emphasizes interruptibility, arguing that prompts should be scheduled according to inferred attentional availability derived from contextual and behavioral cues [6,11]. A third stream, represented by Just-in-Time Adaptive Interventions, frames timing as a dynamic decision based on user state, receptivity, and the likely effectiveness of support at a particular moment [5]. Across these perspectives, a consistent principle emerges: timing is not a neutral delivery parameter but a core component of interaction quality, as it mediates the relationship between system initiative and user cognitive readiness [4–6,11].

The consequences of prompt timing extend beyond immediate usability. Prior research has shown that temporally well-matched interventions are more likely to be interpreted as intelligent, considerate, and trustworthy, whereas poorly timed prompts are often perceived as disruptive even when their content is relevant [3,12]. In voice-based proactive systems, synchronization with the user's ongoing cognitive rhythm has been linked to reduced perceived disruption and more positive evaluations of the system's intelligence [12]. In automotive settings, timing has likewise been treated as a determinant of interaction quality. Studies on in-vehicle auditory–verbal tasks show that opportune moments for interruption can be predicted from driving context and user state [7], while broader reviews of automated vehicle interaction emphasize the importance of adaptive support for situation awareness and trust calibration [8]. Taken together, these studies indicate that

timing plays a constitutive role in shaping both the functional and affective dimensions of human–machine interaction.

Despite this progress, the automotive literature remains largely anchored in operator-centered problems. Much of the existing work addresses warnings, false alarms, takeover requests, and navigation support under conditions in which the user is still primarily understood as a vehicle controller [8,13,14]. This line of research is indispensable, but it does not fully address the logic of interaction in increasingly autonomous, experience-oriented cockpit environments. As automation advances, the user's role shifts from active driving toward supervision, interpretation, and experiential engagement, and the design space for AI prompts expands accordingly. Prior work has also shown that users differ in their preferences for how and when in-vehicle information should be presented [15], reinforcing the point that timing cannot be reduced to a fixed alerting rule. Yet, empirical evidence remains limited on how proactive prompts should be timed when the interaction objective is not only operational support but also cooperative, experience-centered assistance.

The literature, therefore, leaves three unresolved issues. First, timing frameworks are well-developed in mobile, desktop, and adaptive intervention research, but they have not been sufficiently validated in intelligent cockpit contexts, where environmental rhythm, user attention, and AI initiative are tightly coupled [4–6,11]. Second, automotive studies have concentrated on safety-critical or task-control scenarios, leaving the experiential dimension of proactive prompting in autonomous travel underexplored [8,13,14]. Third, few studies isolate timing itself while holding prompt content constant and simultaneously examining its effects across cognitive load, trust, perceived usefulness, and overall experience. This paper addresses that gap by focusing on proactive AI prompt timing in an autonomous tourism-driving scenario, where the user is positioned not as a continuous operator but as an experience-oriented participant. In doing so, it reframes timing as a core variable in cooperative human–machine interaction and develops the argument that the quality of proactive cockpit AI depends on the temporal alignment between system intervention and users' cognitive rhythms.

3. Concept and Hypotheses

3.1. Concept

This study treats prompt timing as a core variable in cooperative human–machine interaction rather than as a secondary property of interface delivery. The conceptual premise is that proactive AI is evaluated not only by the relevance of its communication but also by the temporal relationship between system intervention and the user's cognitive state [3–7]. In intelligent cockpit contexts, this relationship is especially consequential because AI prompts are embedded in an environment characterized by fluctuating attentional demand, changing situational salience, and continuous negotiation between system initiative and user receptivity [2,7,8]. A temporally well-placed prompt may be perceived as intelligent support; the same prompt, delivered at an inopportune moment, may be experienced as interference.

To capture this relationship, the paper introduces temporal alignment as its central conceptual construct. Temporal alignment refers to the degree to which system intervention is synchronized with the user's cognitive rhythm, situational readiness, and momentary capacity to process information. This construct synthesizes three strands of prior work: task-phase models of interruption, which distinguish interventions occurring before, during, and after a task [4]; attention-sensitive alerting and interruptibility models, which emphasize the importance of opportune moments for intervention [6,11]; and adaptive intervention frameworks, which argue that system support should be contingent on user receptivity and contextual appropriateness [5]. In the present study, temporal alignment is not assumed to

be a directly measured psychological mechanism. Rather, it is proposed as the explanatory lens through which the effects of timing on cognitive load, trust, perceived usefulness, and overall experience can be interpreted.

On this basis, prompt timing is operationalized as three temporal positions relative to a salient driving event: Before, During, and After. This task-phase framing was selected because it provides clear temporal delineation while remaining compatible with existing HCI timing research [4,7]. Conceptually, these three positions represent distinct interaction logics. Before prompts are anticipatory: they prepare the user ahead of an event and may enhance predictability and perceived system foresight. During prompts are concurrent: they provide real-time support but may compete with ongoing perceptual and interpretive activity. After-the-prompt retrospectives extend or revisit a completed event but may suffer from reduced situational relevance if they arrive too late to support action. The paper's central proposition is that these temporal positions will not be experienced equally because they imply different degrees of temporal alignment between system behavior and users' cognitive states.

3.2. Hypotheses

Drawing on this framework, three confirmatory hypotheses are developed, providing a structured basis for the study.

H1. *Prompt timing significantly affects perceived cognitive load.*

Prior work on interruption and cognitive processing suggests that system interventions are more demanding when they occur during moments of active attentional allocation than during lower-load or transitional moments [4,6,16]. In the intelligent cockpit, concurrent prompts are therefore expected to impose a greater cognitive burden than anticipatory or retrospective prompts, as they are more likely to compete with ongoing perception and situational interpretation.

H2. *Prompt timing significantly affects trust in the system.*

Trust in automation depends not only on system correctness but also on whether behavior appears reasonable, predictable, and contextually well-judged [3,7,17]. In proactive cockpit interaction, prompt timing may therefore function as a trust cue. Prompts delivered before or at the time of relevance are expected to foster greater trust than delayed prompts because they better signal situational awareness and temporal appropriateness.

H3. *Prompt timing significantly affects perceived usefulness and acceptability.*

The perceived value of a proactive prompt depends on whether users can act on it without experiencing it as intrusive or mistimed [3,5,18]. Accordingly, prompts delivered before or during a relevant event are expected to be evaluated as more useful and acceptable than those delivered after the opportunity for action has diminished.

In addition to these confirmatory hypotheses, the study examines overall user experience satisfaction as a complementary evaluative dimension using the UEQ-S [19]. This dimension is treated as an integrative outcome, rather than as a separate formal hypothesis, because it reflects the combined experiential effect of timing across pragmatic and hedonic qualities.

4. Methods

4.1. Study Design

The study employed a within-subjects experimental design to examine how proactive AI prompt timing affects user experience in an intelligent cockpit scenario. Prompt timing was manipulated at three temporal positions relative to a salient driving event: Before, During, and After.

This task-phase structure follows established HCI timing research and was selected because it provides clear temporal differentiation while remaining operationally tractable in immersive simulation [4]. A within-subjects design was adopted to reduce inter-individual variance and increase statistical sensitivity, given that the research question concerns comparative experience across timing conditions rather than between-user differences.

To control order effects, the three timing conditions were counterbalanced using a Latin square sequence, as shown in Table 1. Each participant encountered all three timing conditions, but in different orders across groups. The design, therefore, isolates timing as the primary manipulated variable while maintaining a comparable interaction structure across conditions.

Table 1. Latin square design for prompt timing conditions.

Group/Condition	Condition 1	Condition 2	Condition 3
A	Before	During	After
B	During	After	Before
C	After	Before	During

4.2. Setup and Participants

The experiment was conducted in an immersive, autonomous, tourism-driving scenario developed in Unity and delivered via a Meta Quest head-mounted display. A VR-based setup was used because the study required both ecological plausibility and precise temporal control over prompt delivery. Prior work has shown that Unity-based immersive environments are well-suited for studying interaction timing because they enable event-triggered manipulation in a realistic yet tightly controllable scenario [20,21].

The scenario simulated a self-driving journey in which participants were not responsible for vehicle control but instead served as experience-oriented passengers and decision-makers. This framing was central to the study's logic: the objective was not to assess driving performance, but to examine how proactive AI prompts are experienced when the user's attention is distributed across observation, interpretation, and lightweight journey decisions. The cockpit interface combined visual and auditory outputs, and the modality was held constant across conditions to ensure that differences in user responses could be attributed to timing rather than presentation format. An overview of the experimental environment is presented in Figure 1.

The sample consisted of 28 university students, mainly enrolled in Master's programs (n = 22), with a smaller group in undergraduate programs (n = 6). Their ages ranged from 22 to 26 years (M = 24, SD = 1.4). The gender split was 19 males and 9 females. This sample size exceeded the minimum requirement estimated in the thesis through a classical power analysis for mean comparison, which indicated a required sample of 25 [22,23]. It also falls within the range commonly reported in prior in-vehicle interaction studies cited in the original thesis [24–27]. Because the design was within-subjects, each participant contributed data for all three timing conditions. All participants were informed about the study's aims and procedures and provided written informed consent prior to participation.

Figure 1. Screenshots of the Unity VR testing environment.

4.3. Experiment Protocol

Regarding the interaction mechanism, participants engaged with the AI prompt system through a coordinated combination of visual and auditory responses and physical input. When a prompt appeared on the central display, participants could view the AI-generated recommendation along with two response options ("Yes" and "No") presented on the interface. Participants were instructed to evaluate the prompt content, reach toward the display to indicate their preferred option, and then confirm their response by pressing the trigger button on the Meta Quest controller. This gesture-and-confirm interaction flow was designed to approximate the tap-to-select behavior expected in production intelligent cockpit systems, while maintaining reliable input registration within the VR environment. Upon confirmation, the driving simulation paused automatically, allowing the participant to remove the headset and complete the questionnaire. After the questionnaire, participants re-entered the VR environment and pressed the trigger to restart the journey. No fixed response-time limit (e.g., a five-second timeout) was imposed on the Yes/No selection. Participants confirmed their responses at their own pace, after which the simulation paused to administer the questionnaire. This was a deliberate design choice: imposing a hard timeout would have introduced an additional, condition-specific time–pressure confound on top of the timing manipulation. Therefore, response latency is not analyzed as a dependent variable in the present study. The vehicle velocity at the prompt-event location was held constant across all three conditions, so that only the temporal placement of the prompt—not its temporal distance to the event nor the driving speed—varied between conditions. Adding response latency and gaze measures is identified as future work in Section 7. Figure 2 illustrates the protocol workflow for user interaction in each experimental condition.

The three experimental conditions were defined as follows: Before, in which the AI provided anticipatory information ahead of the relevant event; During, in which the AI intervened while the event was occurring; and After, in which the AI responded retrospectively after the event had passed; prompt timing was the sole independent variable. The Before prompt concluded at the onset of the event, the During prompt was presented while the event was ongoing, and the After prompt was administered only once the event had concluded. This structure allowed the study to compare distinct temporal positions without altering the prompt's semantic intent.

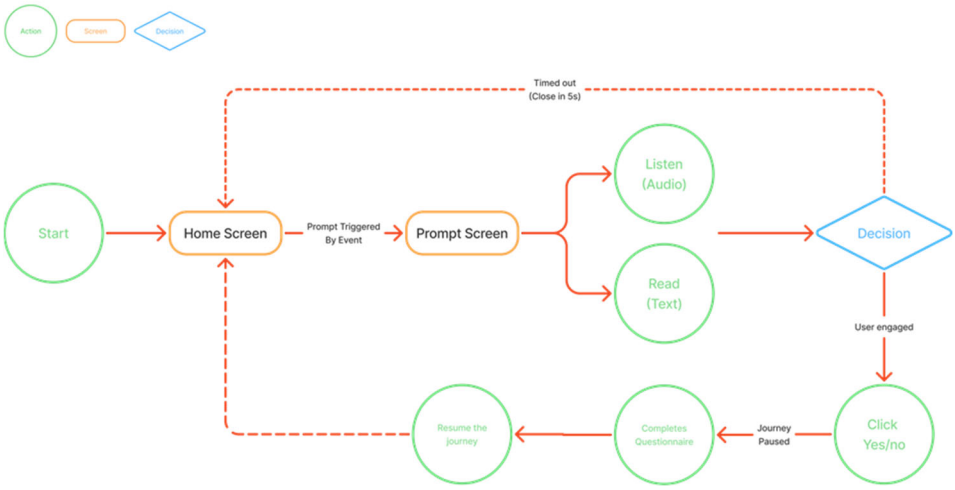

Figure 2. Protocol workflow of user interaction across experimental conditions.

To prevent content effects from confounding timing effects, prompt content was tightly controlled. The study used three categories relevant to autonomous tourism driving: attraction information, commercial recommendations, and safety/fatigue alerts. For each category, semantically equivalent prompts were written for the Before, During, and After conditions. Table S1 in the Supplementary Materials presents the content of each category, organized according to the proposed timing. Thus, the informational goal remained constant while only its temporal placement changed. Prompt texts were also constrained in duration, with auditory delivery limited to approximately 3–8 s to reduce variability from information density or prolonged exposure. This design decision aligns with the study's core premise that timing, rather than content variation, should account for differences in user evaluation.

The experiment unfolded in three stages. First, participants received an introduction to the study purpose and task structure and then entered the VR simulation. Second, they completed the autonomous tourism-driving experience under all three timing conditions. During the journey, the AI presented one prompt from each event category, with timing determined by the assigned Latin square order. Although the study incorporated qualitative feedback, the primary analytical emphasis remained on standardized quantitative measures, consistent with mixed-methods approaches commonly used in interaction-design research [28,29].

Finally, four post-condition instruments were used to evaluate user experience across complementary dimensions. Cognitive load was measured with the NASA-TLX, a widely used subjective workload scale [16]. Trust in the system was measured with the Trust in Automation scale [17]. Perceived usefulness and acceptability were assessed through a 7-point Likert instrument derived from the Technology Acceptance Model [18]. Overall user experience satisfaction was measured with the short form of the User Experience Questionnaire (UEQ-S) [19]. Questionnaire details are provided in Figures S1–S4 of the Supplementary Materials, including the full set of items and measurement scales used in the study. The questionnaire comprised the six NASA-TLX dimensions (mental, physical, and temporal demand, performance, effort, and frustration), the twelve items of the Trust in Automation scale (covering trust, distrust, reliability, predictability, and confidence), six TAM items on perceived usefulness and behavioral acceptability, and the eight UEQ-S bipolar adjective pairs that jointly assess pragmatic and hedonic quality. All items were administered immediately after each timing condition, in the same order across participants, so that responses could be attributed unambiguously to the condition just experienced.

These instruments were chosen because they map directly onto the paper's central evaluative dimensions: mental demand, trustworthiness, perceived value, and integrative experience quality. In addition, semi-structured interviews were used after the experimental session to contextualize the quantitative patterns. The interviews were not treated as a separate confirmatory dataset, but as interpretive support for understanding how participants described the experiential consequences of timing.

4.4. Data Analysis

Each participant completed one questionnaire set after each of the three timing conditions. Questionnaire data were screened for completeness, and all responses were retained. Composite scores were calculated at the scale level. Reverse-coded items were recoded before analysis to ensure directional consistency across measures. All statistical analyses were conducted using IBM SPSS Statistics v. 31 "https://www.ibm.com/products/spss-statistics (accessed on 20 April 2026)"; raw data are provided in Tables S2–S5 whereas detailed outputs are reported in Tables S6–S17 of the Supplementary Materials.

Internal consistency was evaluated for each scale using Cronbach's alpha before inferential testing. The principal inferential analysis used repeated-measures ANOVA to test whether prompt timing significantly affected cognitive load, trust, perceived usefulness/acceptability, and user experience satisfaction. Prior to ANOVA, Mauchly's test of sphericity was performed for each dependent variable. Where the sphericity assumption held, uncorrected degrees of freedom were retained. Pairwise comparisons were conducted using Bonferroni-adjusted post hoc tests. In addition to the omnibus η^2p reported from the repeated-measures ANOVA, two further statistics were computed for the pairwise comparisons to address the recommendations of [30] for small-sample within-subjects designs: (i) the 95% Bonferroni-adjusted confidence interval for each mean difference, and (ii) Cohen's dz, calculated as dz = MeanDiff/SD(diff), where SD(diff) was calculated from the standard error of the difference multiplied by $\sqrt{n}$ with $n = 28$. Reporting CIs and dz alongside η^2p provides a more complete picture of the magnitude, precision, and within-subjects effect size of the observed timing contrasts. This analytical strategy was appropriate for the within-subjects design and for the paper's objective of testing whether temporal position alone produces systematic experiential differences across multiple dimensions.

A further methodological point is that participants' specific Yes/No selections in response to prompts were not treated as behavioral outcome variables in the present study. The interaction was included to create a credible decision context, but the dependent variables were subjective experience measures rather than behavioral choice outcomes. This preserves the paper's focus on how timing shapes experience, trust, and perceived value, rather than on acceptance behavior per se.

5. Results

A total of 84 valid questionnaire responses were obtained from 28 participants across the three timing conditions (Before, During, After), reflecting the within-subjects structure of the study. All questionnaires were completed and retained for analysis. Scale scores were computed at the composite level after reverse coding the relevant items, including the performance dimension of the NASA-TLX and the negatively worded items in the Trust in Automation scale.

Internal consistency was satisfactory to high across all four instruments. Cronbach's alpha was 0.86 for NASA-TLX, 0.88 for Trust in Automation, 0.89 for TAM, and 0.89 for UEQ-S, indicating acceptable reliability for inferential testing. No item removal was required for any scale.

5.1. Descriptive Results

At the descriptive level, the pattern was consistent across most measures. The Before condition produced the highest scores for trust, perceived usefulness/acceptability, and overall user experience satisfaction, whereas the After condition produced the lowest scores on all three measures. By contrast, cognitive load was highest in the During condition and lowest in the After condition. This pattern suggests that anticipatory prompts were generally evaluated more positively, while concurrent prompts were more demanding, and retrospective prompts were less beneficial.

Figure 3 makes three trends visible. NASA-TLX follows an inverted-U pattern, peaking in the During condition and dropping below the Before level in the After condition, indicating that workload tracks attentional competition with the ongoing event rather than the elapsed time since the prompt. Trust in Automation, TAM, and UEQ-S all show a monotonic decline from Before to After, with the steepest drop occurring between During and After, particularly for TAM and UEQ-S. The visual contrast between the inverted-U workload curve and the monotonically decreasing evaluative curves is the first indication that prompt timing affects cognitive cost and experiential value through related yet distinct mechanisms. Table 2 reports the means and standard deviations for each dependent variable across the three timing conditions.

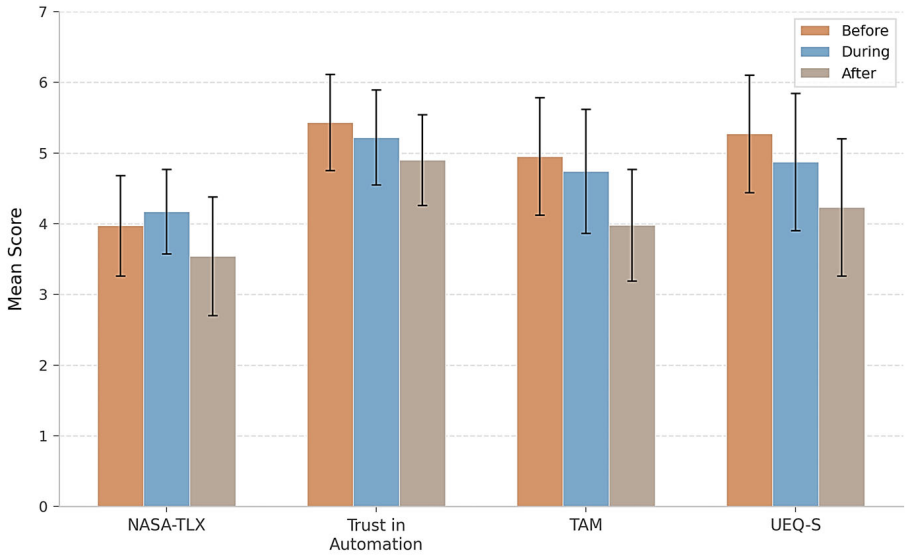

Figure 3. Mean scores (±error bars) for NASA-TLX, Trust in Automation, TAM, and UEQ-S across three conditions (Before, During, After).

Table 2. Standard deviation for each dependent variable across the three timing conditions.

Measure	Before	During	After
NASA-TLX	3.97 ± 0.71	4.17 ± 0.60	3.54 ± 0.84
Trust in Automation	5.43 ± 0.68	5.22 ± 0.67	4.90 ± 0.64
TAM	4.95 ± 0.83	4.74 ± 0.88	3.98 ± 0.79
UEQ-S	5.27 ± 0.83	4.87 ± 0.97	4.23 ± 0.97

Additionally, Mauchly's test of sphericity was performed for each dependent variable prior to conducting the repeated-measures ANOVA. The assumption of sphericity was satisfied in all cases—NASA-TLX ($W = 0.83$, $p = 0.090$), Trust in Automation ($W = 0.98$, $p = 0.715$), TAM ($W = 0.91$, $p = 0.297$), and UEQ-S ($W = 0.96$, $p = 0.595$). Therefore, no corrections were applied, and unadjusted degrees of freedom were used in the main analyses.

5.1.1. Cognitive Load

Prompt timing had a significant effect on perceived cognitive load; $F(2, 54) = 16.78$, $p < 0.001$, $\eta^2p = 0.38$. Bonferroni-adjusted pairwise comparisons showed that the During condition produced significantly higher workload than the After condition (mean difference = 0.63, 95% CI [0.29, 0.96], dz = 0.92, $p < 0.001$), and the Before condition also produced significantly higher workload than the After condition (mean difference = 0.43, 95% CI [0.17, 0.69], dz = 0.81, $p < 0.001$). The difference between During and Before was not statistically significant (mean difference = 0.20, 95% CI [−0.04, 0.44], dz = 0.42, $p = 0.141$). Thus, workload was greatest when prompts occurred during the event and lowest when prompts were delivered after the event.

5.1.2. Trust in Automation

Prompt timing also had a significant effect on trust in the system; $F(2, 54) = 14.74$, $p < 0.001$, $\eta^2p = 0.35$. Pairwise comparisons indicated that trust ratings were significantly higher in the Before condition than in the After condition (mean difference = 0.52, 95% CI [0.28, 0.76], dz = 0.98, $p < 0.001$), and significantly higher in the During condition than in the After condition (mean difference = 0.32, 95% CI [0.05, 0.58], dz = 0.61, $p = 0.016$). The difference between Before and During did not reach statistical significance (mean difference = 0.21, 95% CI [−0.03, 0.44], dz = 0.44, $p = 0.095$). The overall pattern was Before > During > After, with the primary decline occurring for delayed prompts.

5.1.3. Perceived Usefulness and Acceptability

For TAM, the effect of prompt timing was significant and comparatively strong; $F(2, 54) = 38.44$, $p < 0.001$, $\eta^2p = 0.59$. Both the Before and During conditions scored significantly higher than the After condition (Before vs. After: mean difference = 0.96, 95% CI [0.71, 1.22], dz = 1.81, $p < 0.001$; During vs. After: mean difference = 0.76, 95% CI [0.47, 1.05], dz = 1.31, $p < 0.001$). The difference between Before and During was not significant (mean difference = 0.20, 95% CI [−0.13, 0.54], dz = 0.29, $p = 0.402$). The results, therefore, indicate that both anticipatory and concurrent prompts were perceived as more useful and acceptable than retrospective prompts, while the difference between the two timely conditions was limited.

5.1.4. Overall User Experience Satisfaction

Prompt timing significantly affected overall user experience satisfaction as measured by UEQ-S, $F(2, 54) = 33.58$, $p < 0.001$, $\eta^2p = 0.55$. In contrast to the previous measures, all pairwise differences were significant. The Before condition was rated higher than the During condition (mean difference = 0.40, 95% CI [0.10, 0.69], dz = 0.63, $p = 0.006$) and the After condition (mean difference = 1.04, 95% CI [0.69, 1.39], dz = 1.40, $p < 0.001$), while the During condition was also rated higher than the After condition (mean difference = 0.64, 95% CI [0.31, 0.98], dz = 0.93, $p < 0.001$). This produced a clear, ordered pattern of Before > During > After.

5.1.5. Summary

Across all four dependent variables, prompt timing yielded significant main effects with medium-to-large effect sizes. The empirical pattern was not simply that earlier prompts were always better. Rather, the results reveal a more specific contrast between timely prompts (Before and, to a lesser extent, During) and delayed prompts (After). For trust, perceived usefulness, and overall experience, the After condition consistently performed the worst. In terms of cognitive load, the During condition was the most demanding. The only measure on which Before and During differed significantly was UEQ-S, indicating

that anticipatory prompting produced the strongest overall experience advantage, even where differences on trust and usefulness remained directionally rather than statistically distinct. Three patterns are worth noting beyond the headline significances. First, the experiential cost of mistiming is asymmetric: delayed prompts produce the largest decrements in trust, perceived usefulness, and overall experience (dz = 0.98, 1.81, and 1.40 relative to Before, respectively), whereas anticipatory prompts produce the smallest cognitive cost. Second, the contrast between Before and During is consistently the weakest across all four measures ($|dz| \leq 0.63$), suggesting that timely prompts cluster together experientially even when they differ in cognitive demand. Third, UEQ-S is the only measure in which all three pairwise contrasts are significant, indicating that overall experience integrates both the usefulness gain from timely prompts and the workload cost of concurrent ones. Although the η^2p value for NASA-TLX (0.38) is numerically closest to that for trust (0.35), the patterns differ qualitatively: trust, TAM, and UEQ-S decline monotonically from Before to After, whereas workload peaks in the During condition and drops below the Before level in the After condition. The smaller η^2p for NASA-TLX should therefore not be read as a weaker effect of timing, but as a different shape of effect—an inverted-U profile that signals competition for attention rather than a simple decay with delay. Taken together, the results support a graded rather than binary reading of timing effects. Table 3 summarizes the repeated-measures ANOVA results. Table 4 reports the Bonferroni-adjusted pairwise comparisons between time points, including mean differences, Bonferroni-adjusted 95% confidence intervals, and within-subjects effect sizes calculated as Cohen's dz.

Table 3. ANOVA results.

Measure	df	F	*p*	η^2p
NASA-TLX	2, 54	16.78	<0.001	0.38
Trust in Automation	2, 54	14.74	<0.001	0.35
TAM	2, 54	38.44	<0.001	0.59
UEQ-S	2, 54	33.58	<0.001	0.55

Table 4. Bonferroni-adjusted pairwise comparisons.

Measure	Comparison	Mean Diff.	95% CI	*p*	Cohen's dz
NASA-TLX	Before vs. During	−0.20	[−0.44, 0.04]	0.141	0.42
NASA-TLX	Before vs. After	0.43	[0.17, 0.69]	<0.001	0.81
NASA-TLX	During vs. After	0.63	[0.29, 0.96]	<0.001	0.92
Trust in Automation	Before vs. During	0.21	[−0.03, 0.44]	0.095	0.44
Trust in Automation	Before vs. After	0.52	[0.28, 0.76]	<0.001	0.98
Trust in Automation	During vs. After	0.32	[0.05, 0.58]	0.016	0.61
TAM	Before vs. During	0.20	[−0.13, 0.54]	0.402	0.29
TAM	Before vs. After	0.96	[0.71, 1.22]	<0.001	1.81
TAM	During vs. After	0.76	[0.47, 1.05]	<0.001	1.31
UEQ-S	Before vs. During	0.40	[0.10, 0.69]	0.006	0.63
UEQ-S	Before vs. After	1.04	[0.69, 1.39]	<0.001	1.40
UEQ-S	During vs. After	0.64	[0.31, 0.98]	<0.001	0.93

Finally, a correlation analysis was conducted to examine possible relationships among NASA-TLX, Trust, TAM, and UEQ-S within each timing condition. Correlations were calculated separately for each timing condition to avoid collapsing repeated observations from the same participants across conditions. Spearman's rank correlation coefficient was selected, given the relatively small sample size and the questionnaire-based nature of the measures. As reported in Table 5, no correlation reached statistical significance at the 0.05 level. Overall, the observed associations were small to moderate in magnitude, with the largest coefficient emerging between TAM and UEQ-S in the After condition

(ρ = 0.33, p = 0.091). Positive, although non-significant, associations were also observed between Trust and TAM in the Before and During conditions. These findings indicate that the subjective measures obtained pertain to related yet non-redundant facets of the participants' experiences.

Table 5. Spearman's correlations ρ (p) among subjective measures by timing condition.

		Trust in Automation	TAM	UEQ-S
Before	NASA-TLX	−0.15 (0.450)	−0.14 (0.475)	0.14 (0.475)
	Trust in Automation	—	0.27 (0.159)	0.14 (0.489)
	TAM		—	0.27 (0.162)
During	NASA-TLX	0.10 (0.627)	−0.22 (0.266)	0.13 (0.507)
	Trust in Automation	—	0.28 (0.143)	0.18 (0.371)
	TAM		—	0.16 (0.420)
After	NASA-TLX	−0.08 (0.670)	−0.08 (0.703)	0.25 (0.196)
	Trust in Automation	—	0.12 (0.536)	−0.04 (0.830)
	TAM		—	0.33 (0.091)

6. Discussion

This study examined whether the timing of proactive AI prompts influences user experience in an intelligent cockpit scenario for autonomous tourism driving. Across all four dependent variables, prompt timing produced significant main effects, with medium-to-large effect sizes. The overall pattern was consistent but not uniform. Before prompts yielded the highest scores for trust, perceived usefulness/acceptability, and overall user experience satisfaction, whereas After prompts were consistently lowest on these measures. By contrast, During prompts produced the highest cognitive load. The empirical picture is therefore not that one timing strategy dominates on every dimension, but that timing systematically restructures the balance between cognitive demand and positive evaluation. This confirms that prompt timing is not a peripheral delivery parameter; it is a core design variable in proactive cockpit interaction [2–8]. Before turning to the specific patterns, it is important to clarify the analytical role of temporal alignment in what follows. Temporal alignment is introduced not as a directly measured psychological variable but as an explanatory lens that organizes the timing effects observed across the four dependent measures; the construct is used to interpret the empirical pattern, not to claim a separate mechanism that was independently quantified in this study.

A second important result is that the most stable contrast was not among all three conditions, but between timely and delayed prompts. For trust, TAM, and UEQ-S, both Before and During outperformed After, whereas differences between Before and During were generally smaller and statistically nonsignificant except for UEQ-S. This indicates that the principal experiential penalty emerges when the system intervenes too late to be perceived as situationally relevant. In other words, delayed prompting appears to weaken the perceived intelligence and usefulness of the system more consistently than anticipatory prompting improves it. The correlation analysis further indicates that NASA-TLX, Trust in Automation, TAM, and UEQ-S captured related but non-redundant dimensions of participants' experience. No statistically significant correlations emerged within the individual timing conditions, suggesting that the effects of timing should not be interpreted as acting through a single subjective construct. They appear to provide complementary information about participants' responses to the timing of the interaction. Although some positive associations were observed descriptively, particularly between TAM and UEQ-S in the After condition and between Trust in Automation and TAM in the Before and During conditions, these patterns should be interpreted carefully. Overall, the results support the

view that temporal alignment may influence multiple experiential dimensions rather than a single isolated outcome.

The core theoretical contribution of this paper is the proposal of temporal alignment as the organizing concept for interpreting these results. Temporal alignment refers to the degree to which proactive system behavior is synchronized with the user's cognitive rhythm, attentional availability, and situational readiness. The construct is grounded in prior work on interruption timing, opportune moments, and adaptive interventions [4–7,11], but the present study extends that literature by applying it to proactive AI interaction in an intelligent cockpit and by showing that timing effects appear simultaneously across workload, trust, perceived usefulness, and overall experience.

The results support this interpretation in three ways. First, the elevated workload in the During condition suggests that prompts delivered amid ongoing perceptual or interpretive activity create stronger competition for cognitive resources. This is consistent with research on timing, which shows that interventions are more disruptive when they occur outside natural task boundaries or attentional openings [4,6,7]. Second, the trust and TAM results indicate that prompts are not evaluated only on informational grounds. They are also judged as signals of system judgment. When a prompt arrives too late, users appear to infer reduced situational awareness or reduced practical value from the system, even when the content itself remains coherent. Third, the UEQ-S pattern shows that timing differences propagate beyond narrowly functional assessment into broader experiential quality. The fact that Before prompts significantly outperformed During prompts on overall experience, even where trust and TAM differences remained statistically modest, suggests that anticipatory timing contributes not only to utility, but to the felt smoothness and coherence of the interaction.

For intelligent cockpit design, the confirmation of H1 carries a direct implication that should be made explicit. Because During prompts produced the highest cognitive load ($\eta^2p = 0.38$; During vs. After $dz = 0.92$), even modest aggregation of concurrent interventions during dense road segments could compound attentional demand and erode the workload margin that automated driving is intended to provide. In other words, the H1 result is not only consistent with general timing theory; it identifies a specific design risk for cockpit AI that defaults to real-time prompting. Anticipatory prompting therefore emerges in this study not merely as experientially preferable but also as a cockpit-design strategy with a workload-preserving function, particularly valuable in supervisory and experience-oriented automated driving where attentional reserves must be kept available for unexpected situational demands.

Although the present design did not include a direct psychophysiological or behavioral measure of temporal alignment, the construct is empirically anchored in three observable signatures of the data. First, the systematic gradient on cognitive load (During > Before > After) reflects the predicted competition for attentional resources when intervention overlaps ongoing perceptual activity. Second, the asymmetric penalty on trust, usefulness, and overall experience for After prompts maps onto the predicted breakdown of synchronization between system action and user receptivity. Third, the fact that all four measures shift in concert across the three timing conditions—rather than diverging—is consistent with a single underlying alignment dimension expressed across multiple experiential channels. Direct mechanistic validation, for example, through gaze, pupillometry, or continuous workload tracing, remains an important next step; the convergent multi-measure pattern observed here provides a defensible empirical anchor for treating temporal alignment as a useful theoretical construct rather than a purely speculative one, and motivates its use as an organizing frame in cooperative human–machine interaction for intelligent cockpits.

These findings are directly relevant to the Special Issue's emphasis on cooperative human–machine interaction. In conventional interface logic, a prompt is often treated as a discrete unit of information: if the content is correct and relevant, the interaction is presumed to succeed. The present study points to a different conclusion. In proactive AI systems, cooperation depends not only on informational correctness but also on temporal appropriateness. A system that speaks at the wrong moment may be perceived as less intelligent, less trustworthy, and less useful, even if it provides objectively relevant information. This shifts the design problem from content optimization alone toward interaction timing as a condition of cooperation [2,3,8,10].

This shift matters because intelligent cockpits are moving beyond tool-like interaction toward systems that anticipate, recommend, and guide. In that context, prompt timing becomes part of how the AI expresses competence and social legibility. The findings align with prior work suggesting that well-timed proactive dialog is more likely to support trust and acceptance [3,12], while in-vehicle research has already shown that interruptibility depends on situational windows rather than on fixed rules [7]. What this paper adds is an empirical demonstration that in an autonomous tourism-driving context, prompt timing also structures the overall experiential quality of the journey. That contribution is especially relevant for cooperative AI systems designed not only for operational assistance but also for continuous, low-friction human–machine collaboration.

A practical implication of the results is that time should be treated as a design material rather than as a hidden system parameter. The Before condition performed best overall because anticipatory prompts appear to support expectation-setting, perceived system foresight, and smoother integration into the user's activity flow. The During condition retained comparatively high trust and usefulness, but at a higher cognitive cost. This suggests that concurrent prompts remain viable when immediacy is important but should be used selectively and with stronger justification. The After condition, by contrast, consistently underperformed on trust, TAM, and UEQ-S, indicating that retrospective prompting should not be relied upon for primary support where timely action or timely awareness is expected.

These findings support a layered timing logic for intelligent cockpit design. Anticipatory prompts are best suited to previewing upcoming options, clarifying route-relevant opportunities, or supporting lightweight planning before an event occurs. Concurrent prompts should be reserved for cases in which real-time contextual relevance outweighs the cognitive cost of interruption, such as urgent route-linked or safety-relevant information. Retrospective prompts may still be useful, but their role should be reframed toward reflection, summary, or optional follow-up rather than primary intervention. In this sense, the paper's design contribution is not a universal claim that "earlier is always better," but a more precise claim: different prompt timings carry different experiential functions, and delayed prompting is consistently weakest when immediate relevance is central.

More broadly, the results suggest a reframing of proactive intelligence itself. A proactive system should not be understood simply as one that acts first or acts often. It should be understood as one that acts with temporal sensitivity. This interpretation is consistent with definitions of proactive AI that emphasize anticipation and reasoning [1], but it adds a necessary interaction-design qualification: anticipation alone is insufficient if the timing of intervention fails to align with the user's state. In the context of intelligent cockpits, proactive intelligence therefore depends on the ability to balance initiative with restraint, and relevance with timing. From a design standpoint, this means that the question "what should the system say?" cannot be separated from the question "when should the system say it?" [2,3,10].

Lastly, the discussion should be read in light of the study's design scope. The experiment isolated timing while controlling prompt content and modality, thereby strengthening causal interpretation at the level of the manipulated variable. At the same time, the findings apply to the specific context studied here: an immersive, autonomous, tourism-driving scenario evaluated using post-condition subjective measures. The paper's contribution is therefore deliberately bound. It establishes prompt timing as an empirically consequential variable and advances temporal alignment as the conceptual frame through which this consequence can be understood. It does not claim to resolve all timing questions for intelligent cockpits, but it does provide a defensible foundation for treating timing-aware prompting as a central problem in cooperative AI interaction design. A further boundary of the present study concerns the participant sample. The findings were obtained from a relatively small (n = 28) and homogeneous group of university students aged 22–26, who are likely to be more familiar with VR, digital interfaces, and AI-enabled systems than the broader population of future cockpit users. Although this profile is informative for early-stage validation of cooperative AI prototypes—and the within-subjects design reduces inter-individual variance, increasing statistical sensitivity at this sample size—it limits the generalizability of the effects reported here to user groups that may differ on driving experience, familiarity with autonomous systems, age-related attentional dynamics, prior exposure to in-vehicle assistants, and cultural attitudes toward proactive AI. The timing effects reported above should therefore be interpreted as evidence that prompt timing is consequential in this context and for this kind of user, rather than as population-level estimates. Replication with more demographically diverse samples, including older drivers and users with no prior VR exposure, is identified as a priority direction for future work in Section 7.

7. Conclusions

This paper addressed a neglected but consequential question in proactive intelligent cockpit design: when should an AI system prompt? Using a controlled within-subjects experiment in an immersive autonomous tourism-driving scenario, the study showed that prompt timing significantly affects cognitive load, trust in automation, perceived usefulness/acceptability, and overall user experience. The findings were consistent on two points. First, timing is a primary interaction variable, not a secondary delivery detail. Second, the most robust experiential divide is between timely and delayed prompts: prompts delivered before or during relevant events were evaluated more positively than those delivered after those events, while prompts delivered during them also imposed the highest cognitive load.

The paper's main contributions, each anchored directly in the validation results reported in Section 5, are three. First, it provides empirical evidence—convergent across four standard instruments and supported by 95% confidence intervals and within-subjects effect sizes—that prompt timing systematically restructures cognitive load, trust, perceived usefulness, and overall experience in a proactive intelligent cockpit. Second, it identifies an asymmetric experiential profile that has not previously been reported in this combination: delayed prompts carry the largest evaluative penalty (largest dz on trust, TAM, and UEQ-S), while concurrent prompts carry the largest workload cost, with anticipatory prompts dominating on integrative experience. Third, it advances temporal alignment as the conceptual frame that ties these results together and supports a layered design logic in which anticipatory, concurrent, and reflective prompts are assigned distinct cooperative functions. The discussion in Section 6 develops the theoretical and design implications of these contributions; the conclusion records them as the validated core of the paper. In the context of proactive AI, system quality depends not only on whether information is relevant, but on whether intervention is synchronized with the user's cognitive rhythm and situational readiness. This reframes proactive intelligence for cooperative human–machine

interaction: a system is not experienced as intelligent simply because it acts autonomously, but because it acts at a moment that feels appropriate, useful, and cognitively sustainable.

From a design perspective, the study supports a layered timing logic for intelligent cockpit interaction. Before prompts are most effective for anticipatory guidance, expectation-setting, and trust-building. Prompts can provide valuable real-time support, but they should be deployed selectively because of their higher cognitive cost. After prompts appear the weakest as primary interventions and are better suited to reflective or secondary functions. The practical implication is clear: proactive AI systems in intelligent cockpits should move beyond fixed or purely event-triggered prompting toward timing-aware support strategies that treat time as a design material.

The contribution is intentionally bounded. The findings derive from a VR-based, autonomous tourism-driving context with a relatively small, homogeneous sample, and the proposed concept of temporal alignment remains an interpretive construct rather than a directly validated mechanism. Future work should extend this research through field studies with real vehicles, more diverse participant groups, and multimodal or physiological measures that can more directly trace cognitive state. Even with these limits, the present study establishes a firm basis for treating prompt timing as central to the design of cooperative AI systems in intelligent cockpits.

At its core, the paper argues for a simple but consequential design principle: in proactive human–machine interaction, knowing when to speak is part of knowing how to cooperate.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/app16125755/s1, Figure S1: Section A of the Experience Questionnaire—Task Workload; Figure S2: Section B of the Experience Questionnaire—Trust in the AI Assistant; Figure S3: Section C of the Experience Questionnaire—Prompt Helpfulness and Timing; Figure S4: Section D of the Experience Questionnaire—Overall Impression; Table S1: Prompt Types and Content; Table S2: NASA-TLX raw and mean scores; Table S3: Trust in Automation raw and mean scores; Table S4: Trust in Automation raw and mean scores; Table S5: UEQ-S raw and mean scores; Table S6: NASA-TLX Mauchly's Test of Sphericity; Table S7: NASA-TLX Tests of Within-Subjects Effects; Table S8: NASA-TLX Pairwise Comparisons; Table S9: Trust in Automation Mauchly's Test of Sphericity; Table S10: Trust in Automation Tests of Within-Subjects Effects; Table S11: Trust in Automation Pairwise Comparisons; Table S12: TAM Mauchly's Test of Sphericity; Table S13: TAM Tests of Within-Subjects Effects; Table S14: TAM Pairwise Comparisons; Table S15: UEQ-S Mauchly's Test of Sphericity; Table S16: UEQ-S Tests of Within-Subjects Effects; Table S17: UEQ-S Pairwise Comparisons.

Author Contributions: Conceptualization, G.C.; methodology, G.C. and S.P.; validation, G.C.; investigation, S.P.; writing—original draft, S.P.; writing—review and editing, G.C.; supervision, G.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: All subjects gave informed consent for inclusion before participating in the study. This study followed the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of Politecnico di Milano (Application No. 35/2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data supporting the findings of this study are available in the Supplementary Materials. Further inquiries can be sent to the corresponding author.

Acknowledgments: The current study was carried out at Politecnico di Milano's AXD Laboratory (https://bit.ly/axd-polimi-en, accessed on 18 April 2026), in collaboration with the i.Drive Laboratory (http://www.idrive.polimi.it/, accessed on 18 April 2026). Portions of the text in this manuscript were

edited with the assistance of ChatGPT v. 5.4 "https://openai.com/chatgpt (accessed 20 April 2026)" and Grammarly v.1.2.250.1876 "https://www.grammarly.com (accessed 20 April 2026)" to improve clarity and language quality. The authors take full responsibility for the content and its accuracy. The authors gratefully acknowledge Wang Xinyu for his valuable and essential contribution.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Buyukgoz, S.; Grosinger, J.; Chetouani, M.; Saffiotti, A. Two Ways to Make Your Robot Proactive: Reasoning about Human Intentions or Reasoning about Possible Futures. *Front. Robot. AI* **2022**, *9*, 929267. [CrossRef] [PubMed]
2. Gao, F.; Ge, X.; Li, J.; Fan, Y.; Li, Y.; Zhao, R. Intelligent Cockpits for Connected Vehicles: Taxonomy, Architecture, Interaction Technologies, and Future Directions. *Sensors* **2024**, *24*, 5172. [CrossRef]
3. Kraus, M.; Wagner, N.; Minker, W. Effects of Proactive Dialogue Strategies on Human-Computer Trust. In *UMAP '20: Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*; Association for Computing Machinery: New York, NY, USA, 2020; pp. 107–116.
4. Iqbal, S.T.; Bailey, B.P. Effects of Intelligent Notification Management on Users and Their Tasks. In *CHI '08: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*; Association for Computing Machinery: New York, NY, USA, 2008; pp. 93–102.
5. Nahum-Shani, I.; Smith, S.N.; Spring, B.J.; Collins, L.M.; Witkiewitz, K.; Tewari, A.; Murphy, S.A. Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. *Ann. Behav. Med.* **2018**, *52*, 446–462. [CrossRef]
6. Horvitz, E.J.; Jacobs, A.; Hovel, D. Attention-Sensitive Alerting. *arXiv* **2013**, arXiv:1301.6707. [CrossRef]
7. Kim, A.; Choi, W.; Park, J.; Kim, K.; Lee, U. Interrupting Drivers for Interactions: Predicting Opportune Moments for In-Vehicle Proactive Auditory-Verbal Tasks. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2018**, *2*, 175. [CrossRef]
8. Capallera, M.; Angelini, L.; Meteier, Q.; Khaled, O.A.; Mugellini, E. Human-Vehicle Interaction to Support Driver's Situation Awareness in Automated Vehicles: A Systematic Review. *IEEE Trans. Intell. Veh.* **2023**, *8*, 2551–2567. [CrossRef]
9. Ribeiro, M.A.; Gursoy, D.; Chi, O.H. Customer Acceptance of Autonomous Vehicles in Travel and Tourism. *J. Travel Res.* **2022**, *61*, 620–636. [CrossRef]
10. Meurisch, C.; Mihale-Wilson, C.A.; Hawlitschek, A.; Giger, F.; Müller, F.; Hinz, O.; Mühlhäuser, M. Exploring User Expectations of Proactive AI Systems. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2020**, *4*, 146. [CrossRef]
11. Pejovic, V.; Musolesi, M. InterruptMe: Designing Intelligent Prompting Mechanisms for Pervasive Applications. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Seattle, WA, USA, 13–17 September 2014; pp. 897–908.
12. Miksik, O.; Munasinghe, I.; Asensio-Cubero, J.; Bethi, S.R.; Huang, S.-T.; Zylfo, S.; Liu, X.; Nica, T.; Mitrocsak, A.; Mezza, S.; et al. Building Proactive Voice Assistants: When and How (Not) to Interact. *arXiv* **2020**, arXiv:2005.01322. [CrossRef]
13. Azevedo-Sa, H.; Jayaraman, S.K.; Esterwood, C.T.; Yang, X.J.; Robert, L.P.; Tilbury, D.M. Comparing the Effects of False Alarms and Misses on Humans' Trust in (Semi)Autonomous Vehicles. In *HRI '20: Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*; Association for Computing Machinery: New York, NY, USA, 2020; pp. 113–115.
14. Wu, Y.; Yao, X.; Deng, F.; Yuan, X. Effect of Takeover Request Time and Warning Modality on Trust in L3 Automated Driving. *Hum. Factors* **2025**, *67*, 427–444. [CrossRef]
15. Ulahannan, A.; Cain, R.; Thompson, S.; Skrypchuk, L.; Mouzakitis, A.; Jennings, P.; Birrell, S. User Expectations of Partial Driving Automation Capabilities and Their Effect on Information Design Preferences in the Vehicle. *Appl. Ergon.* **2020**, *82*, 102969. [CrossRef] [PubMed]
16. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*; Elsevier: Amsterdam, The Netherlands, 1988; Volume 52, pp. 139–183.
17. Jian, J.-Y.; Bisantz, A.M.; Drury, C.G. Foundations for an Empirically Determined Scale of Trust in Automated Systems. *Int. J. Cogn. Ergon.* **2000**, *4*, 53–71. [CrossRef]
18. Davis, F.D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Q.* **1989**, *13*, 319–340. [CrossRef] [PubMed]
19. Schrepp, M.; Hinderks, A.; Thomaschewski, J. Construction of a Benchmark for the User Experience Questionnaire (UEQ). *Int. J. Interact. Multimed. Artif. Intell.* **2017**, *4*, 40–44. [CrossRef]
20. Shi, X.; Yang, S.; Ye, Z. Development of a Unity–VISSIM Co-Simulation Platform to Study Interactive Driving Behavior. *Systems* **2023**, *11*, 269. [CrossRef]
21. Ugwitz, P.; Šašinková, A.; Šašinka, Č.; Stachoň, Z.; Juřík, V. Toggle Toolkit: A Tool for Conducting Experiments in Unity Virtual Environments. *Behav. Res.* **2021**, *53*, 1581–1591. [CrossRef]

22. Lerman, J. Study Design in Clinical Research: Sample Size Estimation and Power Analysis. *Can. J. Anaesth.* **1996**, *43*, 184–191. [CrossRef] [PubMed]
23. Chow, S.-C.; Shao, J.; Wang, H.; Lokhnygina, Y. *Sample Size Calculations in Clinical Research*, 3rd ed.; Chow, S.-C., Shao, J., Wang, H., Lokhnygina, Y., Eds.; Chapman & Hall/CRC Biostatistics Series; Chapman and Hall/CRC: Boca Raton, FL, USA, 2017.
24. Uang, S.-T.; Hwang, S.-L. Effects on Driving Behavior of Congestion Information and of Scale of In-Vehicle Navigation Systems. *Transp. Res. Part C Emerg. Technol.* **2003**, *11*, 423–438. [CrossRef]
25. Lee, W.-C.; Cheng, B.-W. Effects of Using a Portable Navigation System and Paper Map in Real Driving. *Accid. Anal. Prev.* **2008**, *40*, 303–308. [CrossRef]
26. Lee, W.-C.; Cheng, B.-W. Comparison of Portable and Onboard Navigation System for the Effects in Real Driving. *Saf. Sci.* **2010**, *48*, 1421–1426. [CrossRef]
27. Yang, L.; Bian, Y.; Zhao, X.; Ma, J.; Wu, Y.; Chang, X.; Liu, X. Experimental Research on the Effectiveness of Navigation Prompt Messages Based on a Driving Simulator: A Case Study. *Cogn. Technol. Work* **2021**, *23*, 439–458. [CrossRef]
28. Creswell, J.W.; Plano Clark, V.L. *Designing and Conducting Mixed Methods Research*, 3rd ed.; International Student Edition; Sage: Los Angeles, CA, USA; London, UK; New Delhi, India; Singapore; Washington, DC, USA; Melbourne, Australia, 2018.
29. Preece, J.; Rogers, Y.; Sharp, H. *Interaction Design: Beyond Human-Computer Interaction*, 4th ed.; Wiley: Chichester, UK, 2015.
30. Lakens, D. Calculating and Reporting Effect Sizes to Facilitate Cumulative Science: A Practical Primer for *t*-Tests and ANOVAs. *Front. Psychol.* **2013**, *4*, 863. [CrossRef]